



**Contact**

W. van der Reijden  
Adviseur recordkeeping

T +31 6 55 26 79 52  
wout.van.der.reijden@nationaal  
archief.nl

# memo

## Ontwikkeling metadatamodel eDepot NA

**Datum**

20 oktober 2014

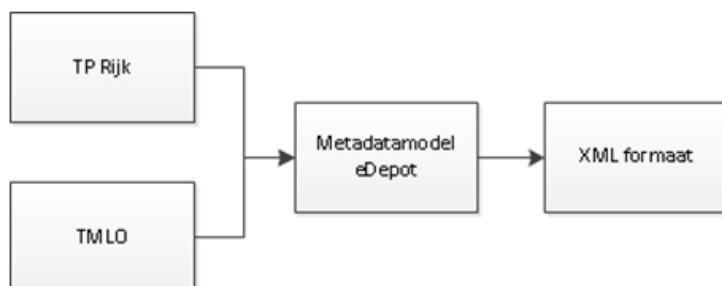
**Auteur**

W. van der Reijden

Onlangs heeft binnen het Nationaal Archief een evaluatie van Medusa, plaatsgevonden. Medusa het XML formaat dat op dit moment tijdens het ingest proces door het NA wordt gebruikt. De Directieraad van het NA heeft op basis van deze evaluatie besloten dat Medusa vervangen dient te worden door een nieuw metadatamodel en een daarvan afgeleid XML formaat. Het model (en XML formaat) is nodig voor het verwerken van de SIP's van de zorgdragers door het NA als onderdeel van het (pre)ingest proces.

De basis voor dit nieuwe model wordt gevormd door twee generieke toepassingsprofielen die op dit moment worden gebruikt door de diverse overheden: het Toepassingsprofiel Metagegevens Rijksoverheid (TP Rijk) en het Toepassingsprofiel Metagegevens Lokale Overheden (TMLO).

Omdat beide profielen niet afzonderlijk voorzien in een XML-formaat neemt het NA de ruimte om deze ontwikkeling zelf tot stand te brengen. Daar het eDepot gebruikt wordt of gaat worden door zowel de Rijksoverheid als de lokale overheden dient het te kunnen omgaan met metadata dat wordt aangeleverd volgens beide generieke profielen.



### Analyse en aanpak

Tussen deze twee profielen zitten een aantal verschillen die door middel van analyse in kaart zijn gebracht.

Tijdens de analyse zijn de twee profielen naast elkaar gelegd en is gekeken waar de verschillen zitten met betrekking tot:

- de elementen (naam)
- het verplichtingsniveau van de elementen
- de herhaalbaarheid van de elementen
- de definities
- de toelichtingen
- waardenverzamelingen en voorbeeldwaarden

In het TP Rijk is ook een kolom aanwezig waarin staat of de waarde van een element automatisch verkregen zou kunnen worden, of niet. Deze kolom is in de analyse buiten beschouwing gelaten, omdat de informatie niet van belang is voor het model en het XML formaat, maar meer dienen als een indicator voor zorgdragers bij het inrichten van een systeem.

De voornaamste conclusies van de analyse:

- a. De elementen komen, op een enkele kleine afwijking in de naam, overeen.
- b. Het verplichtingsniveau van de elementen wijkt bij meerdere (sub)elementen af. Hierbij is in sommige gevallen het TP Rijk strikter, in andere gevallen het TMLO.
- c. De definities wijken enigszins af. Hierbij moet gezegd worden dat het TMLO hier en daar licht afwijkt van de definitie uit de Richtlijn zoals deze is overgenomen in het TP Rijk en het TMLO over het algemeen wel in definities voorziet daar waar de Richtlijn (en dus het TP Rijk) dit niet doet.
- d. De toelichtingen wijken in nagenoeg alle gevallen van elkaar af.
- e. Het TP Rijk voorziet, in tegenstelling tot TMLO, nauwelijks in aanbevelingen voor waardenverzamelingen en voorbeelden van mogelijke waarden.

Voor de ontwikkeling van een vervangend model en afgeleid XML-formaat wordt op de volgende manier met deze verschillen omgegaan:

- Met betrekking tot elementnamen en definities:

Daar waar TMLO een definitie geeft en de TP Rijk niet (het TP Rijk volgt hierin de definities uit de Richtlijn Metagegevens Overheidsinformatie), wordt de definitie van het TMLO overgenomen. Voorbeelden zijn uit de definities verwijderd, om te voorkomen dat er teveel status aan wordt verleend waar dit niet nodig is. Bij afwijkingen in de elementnamen is een keuze gemaakt voor één van de waarden. Daar waar van toepassing zijn tekens als / en () verwijderd omdat zij tot problemen kunnen leiden in namespaces.

- Met betrekking tot verplichtingsniveau en herhaalbaarheid:

Hier is gekozen om de laagste verplichting uit beide profielen aan te houden en

om de herhaalbaarheid van elementwaarden mogelijk te maken als dit in een van beide profielen staat aangegeven.

- Met betrekking tot de toelichtingen:

Hier is gekozen om de toelichtingen ongewijzigd over te nemen, met bronvermelding. Het kan dus voorkomen dat er geen, één of twee toelichting(en) bij de elementen staan.

- Met betrekking tot waardenverzamelingen en voorbeelden:

Er is voor gekozen om de waardenverzamelingen en voorbeelden niet over te nemen in het metadatamodel, omdat deze bij geen van de elementen overeen komen. Een gebruiker van het TP Rijk kan niet worden gedwongen om waarden uit het TMLO te gebruiken en andersom ook niet.

Nu er binnen het NA overeenstemming is over de manier waarop er met de verschillen tussen beide profielen wordt omgegaan, is een eerste versie van het metadatamodel ontwikkeld.

### **Afstemming**

Deze eerste versie van het model wordt gedeeld met de diverse stakeholders via onder meer Breed Netwerk, Pleio en het IMO. Op deze manier kan externe feedback worden meegenomen en waar nodig een aanpassing doorgevoerd. Parallel wordt met de ontwikkeling van het afgeleide XML formaat aangevangen. Als intern model en XML-formaat voor het eDepot zullen beide eenzijdig door het NA worden vastgesteld.

Formele samenvoeging en harmonisatie van de bestaande toepassingsprofielen worden niet gerealiseerd met de vervanging van Medusa. Het vervangingstraject tornt niet aan de vastgestelde profielen, maar voorziet in een lacune aangezien omdat beide profielen niet afzonderlijk voorzien in een XML-formaat. Een dergelijk formaat is naast een model noodzakelijk voor een functionerend eDepot.

De vervanging van Medusa kan wel een katalysator zijn van een formele samenvoeging en harmonisatie van de bestaande toepassingsprofielen.

Gewenste wijzigingen op de bestaande profielen (het kan zijn dat er twijfels bestaan over bijvoorbeeld de herhaalbaarheid van bepaalde elementen, of het verplichtingsniveau) zullen via het regulier beheer ingebracht moeten worden en deze wijzigingen pas kunnen worden doorgevoerd in het metadatamodel en het XML formaat als zij zijn doorgevoerd in één of beide toepassingsprofielen.